

TECHNICAL NOTES¹

General

National estimates of occupational employment in the scientific, technical, and engineering fields for trade and regulated industries² were based on data from the 1994 Occupational Employment Statistics (OES) Survey. The OES Program is a Federal-state cooperative effort in which each state conducts its own survey to produce its estimates. The Bureau of Labor Statistics (BLS) provided each state with survey procedures, technical support, and troubleshooting assistance. The government agencies participating in this program were the 50 State Employment Security Agencies (SESAs) plus the District of Columbia, Puerto Rico, Guam, and American Samoa. For this report, estimates at the national level were produced by BLS-Washington based on data from the fifty states plus the District of Columbia. State-level estimates can be obtained from the individual SESAs.

Scope of Survey

The BLS trade and regulated industries survey covers establishments in SIC codes 40-42, 44-59, and 823. [SIC 823 —“Libraries” —is not reported herein.] The reference dates of this survey were the weeks that included April 12, May 12, or June 12, 1994. The reference date for any particular unit in the survey depended on its SIC code. See the chart below for those SICs covered in this publication..

SIC CODE

REFERENCE DATE

40	May 12
41	May 12
42	May 12
44	Apr 12
45	June 12

1 Portions of the material in these Technical Notes have been excerpted or reproduced from U.S. Department of Labor, Bureau of Labor Statistics: *Occupational Employment Statistics, 1994* (Bulletin 2468, March 1996, “Appendix A. Survey Methods and Reliability of Estimates,” pp. 94-98).

2 A note on industrial classification nomenclature: Industries denoted as “Trade and Regulated Industries” herein are referred to as “Nonmanufacturing Industries” in BLS publications. [See “Scope of Survey” below for listing of industries covered.] In NSF publications “Nonmanufacturing Industries” refer to SIC codes 10, 12-17, 60-65, 67, 70, 72, 73, 75, 76, 78-80 (except 806), 81, 83, 84, 86, 87, and 89. In BLS publications the industries listed above are called “Mining, Construction, Finance, and Services Industries” and are covered by the same survey cycle that NSF reports as “Nonmanufacturing Industries.” In the NSF data brief announcing this publication (“Services Sector S&E Employment Rises, Then Falls Sharply As Engineering and Technician Jobs Are Cut” —NSF 97-322) industries referred to herein as “Trade and Regulated Industries” are termed “Services Industries” to more clearly connote to readers the industries they encompass.

SIC CODE**REFERENCE DATE**

46	June 12
47	Apr 12
48	May 12
49	Apr 12
50	June 12
51	June 12
52	June 12
53	June 12
54	June 12
55	June 12
56	June 12
57	June 12
58	June 12
59	June 12

Method of Collection

Survey schedules were initially mailed out to the personnel offices of almost all sampled establishments. Some of the larger establishments, however, received a personal visit.

Two additional mailings were sent to nonrespondents at approximately six-week intervals. Nonrespondents that were critical to the survey because of their size received a telephone call or a personal visit followup.

Sampling Procedures

The sampling frame for the OES survey was a list of units reported to the state's Unemployment Insurance (U.I.) files. The reference date of the sampling frame was the second quarter of 1993.

Within each state, the universe was stratified by SIC and size class where size class was defined as follows:

Size Class**Employees**

1	1-4
2	5-9
3	10-19
4	20-49
5	50-99
6	100-249
7	250-499
8	500-999
9	1000+

U.I. reporting units with fewer than 5 employees were not sampled in most states; instead, units with 5-9 employees were assigned a larger weight to account for employment in size class 1. U.I. reporting units with 250 or more employees were included in the sample with certainty. The sample sizes needed to calculate state estimates at a targeted relative standard error of 10, 15, or 20 percent for one standard deviation were developed for each SIC across its noncertainty size classes. The sample size for each SIC was determined by calculating averages of occupational rates and averages of coefficients of variation (CVs) for a given set of typical occupations using data from the previous survey round. Within each SIC, the sample size was then allocated proportionally across size classes based on size class employment. The sample was selected systematically with equal probability within each state/(area)³/SIC/size class cell.

The states were given the option of selecting three target relative standard errors in designing their samples. Many states took advantage of this flexibility by varying target relative standard error across SICs in order to balance the cost and reliability of their estimates.

The above allocation resulted in a total initial sample size of 282,861 U.I. reporting units nationally.

Response

Of those sampled, 269,986 were eligible units (i.e., respondents, refusals, unusables, and nonrespondents). Usable responses were obtained from 205,203 units, producing a response rate of 76.0 percent based on units and 73.9 percent based on weighted employment.⁴ See the table below for additional details.

Unit Type	Unit Count	Weighted Empl. Count	
Eligible	269,986	46,214,225	
Respondents		205,203	34,142,013
Nonrespondents		64,783	12,072,212
Out-of-scope and Out-of-business	12,875		
Initial sample size	282,861		

³ Some states opted to further stratify their samples by substate areas.

⁴ Subsequent to the closeout date for national estimates, additional data were collected by the states and used to prepare their respective estimates. Consequently, the response rates in most states were higher than the response rate used to develop national estimates.

Occupational Estimates

Weights were determined for sample units that had usable response. Each weight was composed of two factors, the reciprocal of the probability of selection and a nonresponse adjustment factor (NRAF).

For questionnaires that were not returned or were otherwise unusable, an NRAF was calculated to impute for the missing data. This factor was the ratio

$$\frac{\text{Weighted sample employment of all eligible units in sample}}{\text{Weighted sample employment of all usable eligible units}}.$$

It was calculated for each state/three-digit SIC/size class sampling cell.

The sample employment used to calculate the NRAF was obtained from the sampling frame. If the NRAF in a cell was greater than a predetermined maximum factor (the latter increases as the number of respondents in a cell increases), the cell was collapsed with other homogeneous cells in the industry until the NRAF for the combined cell was not greater than the appropriate maximum factor. If the collapsing procedure terminated (i.e., no more cells were available for collapse) before satisfying the constraint above, then the most recent maximum factor was used. Note that homogeneous cells were adjacent size cells within a state and SIC. The final weight assigned to each usable unit in the sample was the product of the NRAF and the reciprocal of the probability of selection.

A separate ratio estimate of occupational employment was used to develop national estimates. The auxiliary variable used was the 1992 population value of total employment. This variable is also referred to as cell benchmark employment, denoted by M_{ij} . The term

$$\frac{M_{ij}}{\sum_k W_{ijk} E_{ijk}}$$

is known as the benchmark factor. It is the ratio of cell benchmark employment to cell weighted reported total employment. The estimation formula below produced final estimates (\hat{P}_{ij}) of occupational employment through benchmarking, that is, the process of multiplying the cell's weighted reported occupational employment ($\sum W_{ijk} P_{ijk}$) by its benchmark factor.

$$\hat{P}_{ij} = \left(\sum_k W_{ijk} P_{ijk} \right) \left(\frac{M_{ij}}{\sum_k W_{ijk} E_{ijk}} \right)$$

where \hat{P}_{ij} = estimated employment for occupation P in industry i and size class j

i = a three-digit industry

j = size class

k = establishment

W_{ijk} = weight for establishment k in industry i and size class j after adjusting for nonresponse

P_{ijk} = reported employment for occupation P in establishment k within industry i and size class j

E_{ijk} = reported total employment for establishment k in industry i and size class j

M_{ij} = population value of total employment for industry i and size class j.

The estimated employment for an occupation at the three-digit industry i level was obtained by summing the occupational employment estimates P_{ij} across all size levels j within industry i.

$$\hat{P}_i = \sum_{j=1}^{L_i} \hat{P}_{ij}$$

where L_i was the number of size levels j in industry i.

Similarly, the estimated employment for an occupation at the two-digit industry g level was obtained by summing the occupational employment estimates P_i across all three-digit industries i within two-digit industry g.

$$\hat{P}_g = \sum_{i=1}^{L_g} \hat{P}_i$$

where L_g was the number of three-digit industries i in industry g.

It is important to note, however, that because of publishability requirements, rounding adjustments were made such that occupational employment estimates at the three-digit industry level may not sum to the two-digit level estimates.

Variance Estimates

Estimates of sampling error were calculated on survey estimates to allow users to determine whether or not the occupational estimates were reliable enough for their needs. Only a probability sample can be used to estimate sampling error from a sample.

The formulas used to estimate the variance, a common measure of sampling error, were based on the sample design and on the method of estimation. In the OES survey, the formula used to estimate the variance of occupational employment was a subsample replication technique called the jackknife random group. The jackknife derives R estimates of total occupational employment from R subsamples of the parent sample by excluding one random group at a time. The jackknife then estimates the variance of the parent sample's employment estimator from the variability between the R employment estimates.

The variance for an occupational employment estimate at the three-digit industry i/size class j level is

$$S^2(\hat{P}_{ij}) = \frac{1}{(R)(R-1)} \sum_{r=1}^R (\hat{P}_{ijr} - \hat{P}_{ij})^2$$

Where

$$\begin{aligned} S^2(\hat{P}_{ij}) &= \text{estimated variance of } \hat{P}_{ij} \\ R &= \text{number of random groups} \\ \hat{P}_{ij} &= \text{estimated employment for occupation P in industry i and size class j} \\ \hat{P}_{ijr} &= \text{estimated employment for occupation P in industry i, size class j, and} \\ &\quad \text{and subsample r} \\ \hat{P}_{ij} &= \text{estimated mean employment for occupation P in industry i and size class j} \\ &\quad \text{across R subsamples} \end{aligned}$$

The above formula for variance has been simplified. The actual formula includes corrections for finite populations.

The variance for an occupational employment estimate at the three-digit industry i level is obtained by summing the variances $S^2(\hat{P}_{ij})$ across all size levels j within industry i.

$$S^2(\hat{P}_i) = \sum_{j=1}^{L_i} S^2(\hat{P}_{ij})$$

where L_i is the number of size levels j in industry i.

Similarly, the variance for an occupational employment estimate at the two-digit industry g level is obtained by summing the variances $S^2(\hat{P}_i)$ across all three-digit industries i within industry g.

$$S^2(\hat{P}_g) = \sum_{i=1}^{L_g} S^2(\hat{P}_i)$$

where L_g is the number of three-digit industries i in industry g.

Reliability of Estimates

Estimates developed from the sample may differ from the results of a complete census of all the establishments in the sampling frame. Two types of error, sampling and nonsampling, are possible in an estimate based on a sample survey. Sampling error occurs because observations are made only on a sample, not on the entire population. Nonsampling error can be attributed to many sources, e.g., an inability to obtain information about all cases in the sample, differences in the respondents' interpretation of questions, inability or unwillingness of respondents to provide correct information, errors made in recording, coding, or processing the data, errors made in estimating values for missing data, and failure to represent all units in the population.

Sampling error arises because the particular sample used in this survey is only one of a large number of possible samples of the same size that could have been selected with the same sample design. Estimates derived from those different samples would differ simply as a result of random effects. Relative standard standard errors that are a measure of that sampling error effect are presented in this publication. The relative standard errors of a survey estimate measure the variation among the estimates from all possible samples. The relative standard error is the standard error of the estimate divided by the employment estimate for that occupation. Thus, it shows the size of the standard error relative to the occupational estimate itself.

Use of the relative standard error enables the analyst to construct a confidence interval around the occupational estimate. The confidence interval includes the average value of the estimates obtained from all possible samples (of that size and design) at a confidence level specified by the analyst. If no nonsampling error is present (which is unlikely) the *a priori* interval will contain the true value with the confidence level specified.

To construct the confidence interval, divide the relative error shown in the table by 100 and multiply the result by the occupational estimate. The confidence interval is the occupational estimate, plus or minus the number resulting from the calculation described above. This estimate yields a confidence level of approximately 68 percent. That is, in 68 percent of the possible samples, the "true value" (neglecting nonsampling error) will be contained in the interval. Most analysts prefer to have a confidence level higher than 68 percent. If a 90 percent confidence level is desired, multiply the number produced from the calculation in the first sentence above by 1.6. For a 95% confidence level, multiply by 1.96. For almost full confidence (99 percent), multiply by 2.57.

For example, suppose the occupational employment estimate for chemist is 5,000 with an associated relative standard error shown on the table of three. The 68 percent confidence interval will then be $(3/100) \times 5,000$ or the chemist estimate plus or minus 150. The "true value" will be contained in the interval of 4,850 to 5,150 about 68 percent of the time. For 95 percent confidence, multiply 150 times 1.6 = 240. The 95 percent confidence interval is 4,760 to 5,240. It is important to remember that nonsampling error can have important effects on the accuracy of the estimate. Unfortunately nonsampling errors can be very difficult to measure and are not available.

The relative standard errors primarily indicate the magnitude of the sampling error. They do not measure nonsampling error, including any biases in the data. Many edit and quality control procedures are used to reduce the nonsampling error caused by mistakes in recording, coding, and processing the data. The adjustments made for nonrespondents assumed that the characteristics of the nonrespondents are the same as those of the respondents at a given level. To the extent that this is not true, bias is introduced in the data. The magnitude of this bias is not known.

Particular care should be exercised in the interpretation of small estimates or small differences between estimates, because of relatively large sampling errors and the unknown magnitude of the biases.

Quality Control Measures

As described above, the OES Survey is a Federal-state cooperative effort in which states conduct their own surveys.

A major concern with a cooperative program like OES is accommodating state-specific publication needs with limited resources while standardizing survey procedures across all fifty states and the District of Columbia in order to produce quality estimates. The control on sources of nonsampling error in this decentralized environment can be particularly difficult. In addition, the review and validation function is spread across eight regional offices, thus leading to procedural differences between regions. Examples of quality control measures employed by BLS are the Survey Processing and Management System (SPAM) and the Estimates Delivery System (EDS), which were developed to provide a consistent and automated framework for state procedures and to reduce the workload at state, regional, and national levels.

By standardizing data processing activities (i.e., validating the sample frame, allocating and selecting the sample, refining mailing addresses, addressing envelopes and mailers, editing and updating questionnaires, producing management reports, and producing estimates) across all states, the use of SPAM and EDS has also standardized the survey methodology. This has significantly reduced the number of errors on the data files as well as the time needed to review them.

Other quality control measures used in the OES survey include

- followup solicitations of nonrespondents (especially essential nonrespondents),
- review of schedules to verify the accuracy and reasonableness of the reported data,
- adjustments of atypical data reporters,
- validation of the nonresponse adjustment factors,
- validation of the benchmark employment figures and of the benchmark factors, and
- validation of the analytical tables of estimates (at the two and three-digit SIC levels).